

Received October 23, 2020, accepted November 21, 2020, date of publication November 25, 2020, date of current version December 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3040426

Wafer Edge Yield Prediction Using a Combined Long Short-Term Memory and Feed-Forward Neural Network Model for Semiconductor Manufacturing

DASOL KIM^{1,2}, MINTAE KIM¹, AND WOJU KIM¹

¹Department of Industrial Engineering, Yonsei University, Seoul 03722, South Korea

²Samsung Electronics Company Ltd., Gyeonggi-do 18448, South Korea

Corresponding author: Wooju Kim (wkim@yonsei.ac.kr)

This work was supported in part by the Korea Ministry of Land, Infrastructure and Transport (MOLIT) as [Innovative Talent Education Program for Smart City] and in part by Samsung Electronics Co., Ltd.

ABSTRACT In semiconductor manufacturing, maintaining a high yield and ensuring accurate yield prediction are considerably important for improving productivity, customer satisfaction, and enhancing profitability. Despite its importance and merits, achieving wafer yield prediction with high quality and accuracy is challenging. In this paper, we propose a method for wafer edge yield prediction using a combined long short-term memory (LSTM) and feed-forward neural network (FFNN) model. Unlike previous research, we focus on the edge yield because of the higher yield loss at the wafer edge. The combined LSTM-FFNN model uses a dataset divided into two types according to data characteristics. Time-series data are used in the case of LSTM, and non-time-series data are fed into the FFNN. When preparing the time-series data, comprising data related to the equipment and chambers, data of different chambers do not overlap, thereby rendering them as independent entities. The proposed model outperforms other models in terms of all evaluation metrics. The coefficient of determination of the proposed combined LSTM-FFNN model is 34.14%, which is almost 13% higher than that of the other compared models on average.

INDEX TERMS Feed-forward neural networks, long short-term memory, semiconductor manufacturing, wafer yield prediction.

I. INTRODUCTION

In recent years, as advanced technologies such as smartphones, deep learning, the Internet of Things, and artificial intelligence have emerged, the demand for semiconductors has increased exponentially. Meanwhile, semiconductor manufacturing, which involves several process steps, is becoming increasingly complex and difficult to manage. The semiconductor manufacturing process involves monitoring numerous parameters from the early stages of production up to the packaging of an end product [1]. Metrology, the most important parameter among these, is the key to achieving high product quality in semiconductor manufacturing. In general, metrology is measured physically using a helium ion microscope or scanning electron microscope (SEM) [2], [3]. Each wafer is measured after each process; thus, the quality of each wafer can be estimated. However, this is

impractical because every measuring process added between each pair of contiguous processes can significantly increase the total time of production [4]. Combined focused ion beam and SEM could be a possibility for on-line metrology [5].

Therefore, virtual metrology, one of many fabrication parameters, has been developed to augment physical metrology. Recently, virtual metrology has been employed for obtaining additional information from the analysis of scrapped wafers or electrical test values. Virtual metrology, which significantly enhances fabrication productivity and quality assurance [6], is a parameter that correlates various sensors on the process equipment with metrology. Because valuable data can be obtained without wafers having to proceed through metrology steps, it is possible to reduce the cost of metrology tools and the overall process time [7]. Despite the advantages of virtual metrology, checking wafer quality should not be solely dependent on it because of the need for consistency, which necessitates periodic updates. Both metrology and virtual metrology entail time-series data.

The associate editor coordinating the review of this manuscript and approving it for publication was Rahul A. Trivedi¹.

Certain machine parts that constitute the process equipment are worn away gradually through repetitive processes, and these machine parts and materials such as photoresist, etching gas, and other chemicals should be periodically replaced to maintain high quality. Engineers monitor the fabrication parameters in real time for each equipment chamber independently. To achieve the target value, Run-to-Run (R2R) [8] control is employed, and engineers refer to previous parameter values and intentionally tune process settings, such as temperature, pressure, and gas flow, for the next parameters, thereby creating time-series data.

The semiconductor wafer yield is defined as the ratio of the number of good chips to the total number of chips. Yield is a widely used performance metric in semiconductor manufacturing; moreover, maintaining a high yield via reliable and accurate quality control is a key performance metric. Accurate yield prediction is highly important for improving productivity, customer satisfaction, and enhancing profitability [9]. Reference to yield predictions enables semiconductor manufacturers to implement supply chain management and guarantee high-quality products. Yield prediction is becoming more important as a future task [10]. Despite its importance and merits, wafer yield prediction has a significantly challenging goal of being systematic with high quality and accuracy. Under these conditions, many engineers in semiconductor manufacturing have attempted to predict the yield constantly in practice. These yield prediction models still need to resolve poor performance and are one of the most important goals for the department of yield enhancement. However, yield prediction models focus on the total average wafer yield. Moreover, the wafer yield differs for different regions of the wafer, consisting of the inner and edge yields. In general, the edge yield is substantially lower than the inner yield. Although the yield at the edge only accounts for a small proportion of the wafer, it has a significant influence over the decrease in the total yield. Thus, the edge yield is an important aspect of wafer yield that should be prioritized. A typical example of the inner and edge areas of a wafer is illustrated in Fig. 1.

The process variation across a wafer may be greater at the edge compared to that at the center, resulting in a higher yield loss at the wafer edge [11]. The thinning process could induce damage at the wafer edge, which would directly impact the physical yield [12]. According to Yavas, several factors can lead to significant edge yield loss. These factors include non-uniformities in the wafer thickness and etch profiles due to plasma inhomogeneity toward the wafer edge, wafer bow due to film stress, residues at the bevel and backside, chuck damage by reactive gases or particles, and plasma or handling-induced mechanical damage at the bevel [13]. Almost every process is focused on the inner area of a wafer, which may explain the aforementioned results. Because process variation at the edge contributes to edge yield loss, the edge yield should receive more attention. This motivated us to propose a wafer edge yield prediction model using edge parameters.

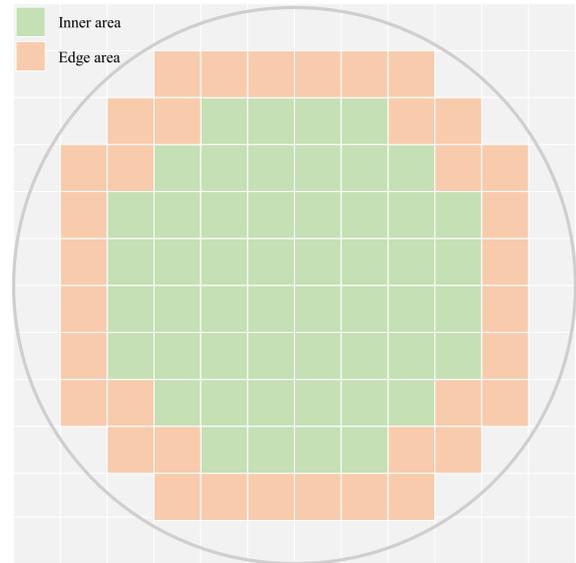


FIGURE 1. Example showing two different wafer areas: the inner and edge areas.

In this study, we used several fabrication parameters, that is, metrology, virtual metrology, and equipment information obtained during critical steps, as the features. We additionally introduced a wafer edge yield prediction model that combines long short-term memory (LSTM) with a feed-forward neural network (FFNN) model. The main contributions of this work are summarized as follows:

- We divide the dataset and input the smaller datasets into the corresponding models to understand the characteristics of the dataset. Because the data in semiconductor manufacturing evidently exhibit the time-series property by R2R control or engineers' tuning to achieve the target value, this characteristic should be included in the wafer yield prediction model. Sequential data are used in the case of LSTM, and the remaining data are fed into the FFNN.
- We use a time-series dataset consisting of data related to the equipment and chambers. When several chambers are connected to a single piece of equipment, they should be regarded as independent entities because they are controlled by different recipes. Data of different chambers do not overlap with each other when this sequential dataset is prepared for the LSTM model.
- We propose a method for wafer edge yield prediction, which is the key to enhancing edge yield, as opposed to focusing on predicting the total yield of the wafer.

The remainder of this paper is organized as follows. In Section 2, we describe previous studies in detail. In Sections 3 and 4, we present the proposed method and compare our experimental results with those of other well-known methods. Finally, we present a brief conclusion and discuss future work in Section 5.

II. RELATED WORK

Several prior studies have focused on wafer yield prediction, and the results of sequential research have been used in semiconductor manufacturing.

A. PRELIMINARY YIELD PREDICTION EXPERIMENTS

Shin *et al.* proposed a hybrid machine-learning strategy using fabrication parameters. Their strategy involved a neural network and memory-based learning for lot-based yield prediction in semiconductor manufacturing [1]. Li *et al.* proposed a comprehensive data mining method for predicting and classifying product yields in semiconductor manufacturing processes using a genetic programming approach [14]. Chien *et al.* suggested a framework for yield prediction and the classification of abnormal process stages in semiconductor manufacturing using the Kruskal–Wallis test and a decision tree [15]. However, these lot-based prediction models have limitations because only sampling and measuring two to three wafers in one lot that includes 24 wafers may entail an excessive assumption. This ultimately led to the proposal of wafer-based yield prediction models.

Nam *et al.* proposed a prediction model to predict wafer yield based on virtual metrology process parameters in semiconductor manufacturing [16]. Chien *et al.* proposed a novel data-driven approach to analyze big data generated during semiconductor manufacturing. The method is intended for low-yield diagnosis to detect the root causes of processes for yield enhancement. The data reflect the production process steps, tools, recipes, and vendors [17]. Jang *et al.* proposed a novel yield prediction model based on deep neural network algorithms by using the spatial relationships among the positions of dies on a wafer and die-level yield variations extracted from a wafer test without process parameters [18]. Although they used wafers for yield prediction, their approach is problematic in that metrology parameters are not included as features. Metrology parameters are the most powerful features for predicting wafer yield with high reliability because these parameters are the only values that result from actual measurements. However, the use of metrology for yield prediction is difficult because values will inevitably be missing from metrology owing to productivity and time limitations.

An *et al.* suggested an efficient way to distinguish high yield and low yield using a stepwise support vector machine. Measurements of the unit voltage, current, and other electrical characteristics were used for yield prediction after fab-out [19]. However, real-time prediction at the fabrication level is not possible because the input dataset is obtained after fab-out.

B. SEQUENTIAL RESEARCH IN SEMICONDUCTOR MANUFACTURING

Yang *et al.* proposed a novel approach that incorporates the interactions among spec-out events using spec-out event network analyses with time-series process sensor data such as temperature, pressure, and voltage data [20]. Lee *et al.* proposed a convolutional neural network (CNN) model, in which a receptive field tailored to multivariate sensor signals slides along the time axis, to extract fault features. In semiconductor manufacturing processes, all recipe parameters should reach their individual set points in a timely

manner and maintain the set points without severe fluctuations for specified process durations [21]. Chen *et al.* proposed a method for anomaly detection in semiconductor manufacturing through time-series forecasting using three models: autoregressive integrated moving average, multi-layer perceptron, and LSTM [22]. Kim *et al.* proposed fault detection and diagnosis using self-attentive CNNs for variable-length sensor data in semiconductor manufacturing [23].

III. PROPOSED METHOD

In this section, we present the details of the proposed model for wafer edge yield prediction. First, various input features for the proposed model are described. Second, we provide a detailed account of the combined LSTM–FFNN model and its ability to effectively use both time-series data and non-time-series data.

A. DESCRIPTION OF INPUT FEATURES

The input features are summarized in Table 1. Four types of input features are used: metrology, virtual metrology, equipment output data, and equipment information. Furthermore, semiconductor manufacturing involves numerous parameters; therefore, the selection of an appropriate and optimal dataset is necessary. Among the several selection methods, domain knowledge from practical experience and statistical analysis is adopted to select the features. While engineers focus on yield analyses for defining the root cause, they not only identify the critical process steps but also the fabrication parameters when the root cause matches the specific fabrication parameters in the critical process steps. Three critical steps, namely A, B, and C, from among hundreds of process steps, and three additional steps, namely D, E, and F, determined via statistical analysis (Kruskal–Wallis), were selected [24]. The Kruskal–Wallis test entails the non-parametric analysis of variance to compare several independent samples. The results of the test are summarized in Table 2. We represent the edge yields as Y (numerical values) and all process steps as X (categorical values). Among all the process steps, three steps, namely D, E, and F, were found to have the lowest p-values. Steps B and A, which were selected based on domain knowledge, are in the fourth and fifth places on the list, respectively.

Step A is the most critical step; this is the reason metrology and virtual metrology data are obtained after step A has been completed. In the processing of step A, it is found that the edge pattern is slightly different from the center pattern because of the wafer topography. This difference is quantified by metrology and virtual metrology, and also affects the edge yield directly. Predicting wafer edge topography and then interpreting this to yield and yield degradation classification might be a more accurate method to determine edge yield, however checking wafer topography is costly in terms of money and time. Thus, in this study we predict edge yield directly. Because the proposed prediction model is for wafer edge yield, we only selected six metrology and twenty-three

TABLE 1. Input features.

Feature	Counts	Description	Wafer area	Related process	Measuring rate
Metrology	6	In-line measurement	Wafer edge	A	Under 100%
Virtual metrology	23	Calculated values by pre-trained algorithm	Wafer edge	A	100%
Equipment output data	3	Time value from process equipment	Wafer average	A,B,C	100%
Equipment information	139	Names of the equipment in the six process steps	-	A,B,C,D,E,F	100%

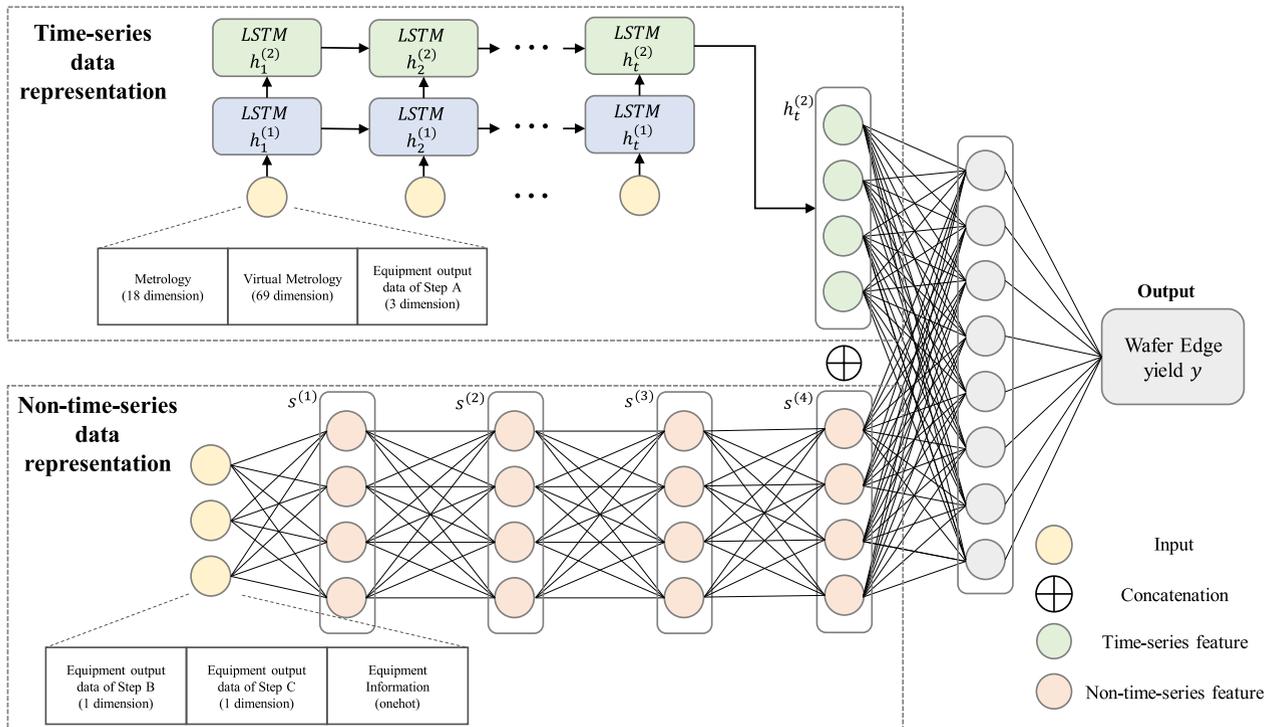


FIGURE 2. Architecture of the proposed method.

TABLE 2. Result of the kruskal-wallis test.

Y(numerical)	X(categorical)	p-value
Edge yield	STEP D	5.20E-93
	STEP E	2.54E-35
	STEP F	1.32E-33
	STEP B	5.65E-31
	STEP A	8.01E-27

virtual metrology data corresponding to the wafer edge area among tens of metrology and virtual metrology. All of the virtual metrology used corresponds to data for the additional values that make input features informative with metrology. The measuring rate for metrology is less than 100%. The output data obtained regarding the equipment are plasma-on time values obtained from the process equipment, and the equipment is set to plasma-on by each chamber after product maintenance and plasma-off after the determined time passes or serious events occur. These time values verified that the start and end times affect the wafer yield due to fluctuations in the processing rate. We extract these data from steps A, B, and C. There are also hundreds of other equipment output data. Some of them were used for input features, but they did

not perform well because several wafers exhibited the same value for each feature. Finally, equipment name information was obtained for all six steps and converted to a one-hot encoded dataset.

B. COMBINED LSTM-FFNN MODEL

In this subsection, we discuss the proposed prediction model that combines LSTM with the FFNN. The architecture of the proposed model is illustrated in Fig. 2.

In general, the wafer yield prediction model uses only the non-time-series data of each process step. However, certain data clearly have time-series characteristics. Each data value is connected to another because engineers in semiconductor manufacturing refer to previous parameter values and engage in a series of fine-tuned value adjustments for the next set of data, thereby creating time-series data. Therefore, time-series data should be taken into account when using fabrication parameters as input features. The combined LSTM-FFNN model proposed in this paper effectively uses both time-series data and non-time-series data to improve the yield prediction performance.

First, we extract the features from the time-series data obtained via metrology, virtual metrology, and equipment output data corresponding to step A using the LSTM architecture. We employ multi-layer LSTM as networks with stacks of several LSTM models, where the hidden representation of the previous layer is used as the input for the next layer. Stacked-LSTM can solve more complex problems and extract hidden hierarchical information. Assuming that l denotes a layer, the hidden state of time-step t in layer $h_t^{(l)}$ can be calculated as follows:

$$h_t^{(l)} = LSTM(x_t, h_{t-1}^{(l)}), \quad (1)$$

where x_t denotes the input metrology data at step t .

In this study, we use a two-layer stacked-LSTM model, and the final encoded metrology feature can be expressed as $h_t^{(2)}$.

Second, we extract the features from non-time-series-data, that is, the equipment output data of steps B and C and the one-hot encoded equipment name information. To this end, we used multilayer neural networks. The output of the l -th layer can be calculated as follows:

$$s^{(l)} = \sigma(W^{(l)}X^{(l)} + b^{(l)}), \quad (2)$$

where $X^{(l)}$, $W^{(l)}$, and $b^{(l)}$ are the input, weight, and bias of the l -th layer, respectively, and σ is the activation function. In our work, we used the rectified linear unit (ReLU) as the activation function [25].

We designed fully connected neural networks with four hidden layers. Each hidden layer consisted of 128 nodes. The final encoded feature of the non-time-series data can be expressed as $s^{(4)}$.

Finally, we concatenate the features $h_t^{(2)}$ and $s^{(4)}$, which represent time-series data and non-time-series data, respectively, and the final wafer yield y_i is obtained using a fully connected neural network with one hidden layer.

The network is trained by conducting back-propagation using the Adam optimizer with a learning rate of 0.001. We use the mean squared error (MSE) as the loss function, which can be calculated as follows:

$$\min_{\theta} = \frac{1}{N} \sum_i^N (f(\theta; x_i) - y_i)^2, \quad (3)$$

where $\{x_i\}_{i=1}^n$ are the training inputs, $\{y_i\}_{i=1}^n$ are the wafer yield labels, θ are the weights of our architecture, and f is the prediction function in our proposed combined LSTM-FFNN model.

IV. EXPERIMENT

A. EVALUATION DATASET

The dataset we used consisted of data regarding an advanced 3D vertical-NAND flash memory device from a semiconductor manufacturing company in South Korea. Both the product name and process step information are kept confidential for security reasons. Data relating to a total of 89,093 wafers were collected. In addition, we used a time-series dataset with

data from equipment and chambers, and nearly 70 equipment chambers corresponding to the process step A were used. The dataset with time-series characteristics is prepared using a three-sequence length for each equipment chamber. A sequence length of 3 is used because engineers usually tune the recipe for the equipment chamber by referring to 3 points of change when monitoring fabrication parameters. An overview of the dataset is illustrated in Fig. 3. The training set was composed of 71,108 data samples, and the test set consisted of 17,778 data samples.

B. EVALUATION METRIC

To objectively evaluate the performance of the model, three evaluation metrics were adopted to compare the quality of different models: coefficient of determination (R^2), MSE, and mean absolute error (MAE). These metrics are mainly used to evaluate the performance of the regression model.

$$MAE = \frac{1}{N} \sum_i^N |y_i - \hat{y}|, \quad (4)$$

$$MSE = \frac{1}{N} \sum_i^N (y_i - \hat{y})^2, \quad (5)$$

$$R^2 = \frac{\sum_i (y_i - \hat{y})^2}{\sum_i (y_i - \bar{y})^2}, \quad (6)$$

where y_i , \hat{y} , and \bar{y} are the actual value of y , the predicted value of y , and the mean value of y , respectively. N denotes the number of observations.

C. COMPARING METHODS

In our experiments, we used four regression algorithms to build the yield prediction models: neural networks, support vector regression, decision tree, and partial least square (PLS) regression.

- Neural networks [26] are widely used computing systems inspired by biological neural networks for time-series prediction, nonlinear multivariate prediction, and anomaly detection in the field of manufacturing. We used feed-forward neural networks with four hidden layers and the Adam optimizer; the MSE loss function is the same as that employed in the combined LSTM-FFNN model. The only difference between this neural network and our model is the formation of the inputs. Neural networks receive their inputs without taking the sequence of the dataset into account, which enables us to verify the effectiveness of a time-series dataset in semiconductor manufacturing.
- Support vector regression [27] is a regression algorithm; it adds an ϵ -insensitive loss function for solving regression problems via the support vector machine [28], used for solving classification problems. The support vector algorithm is advantageous for complex models, and it is sufficiently simple for analyzing a space-related nonlinear problem mathematically. This is because it corre-

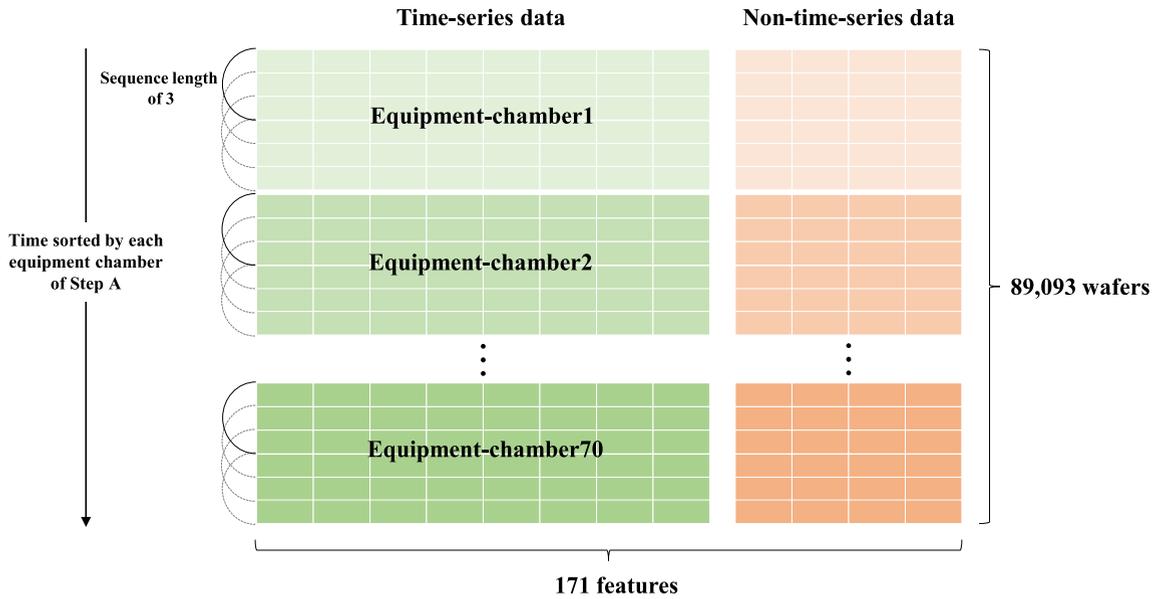


FIGURE 3. Overview of the dataset.

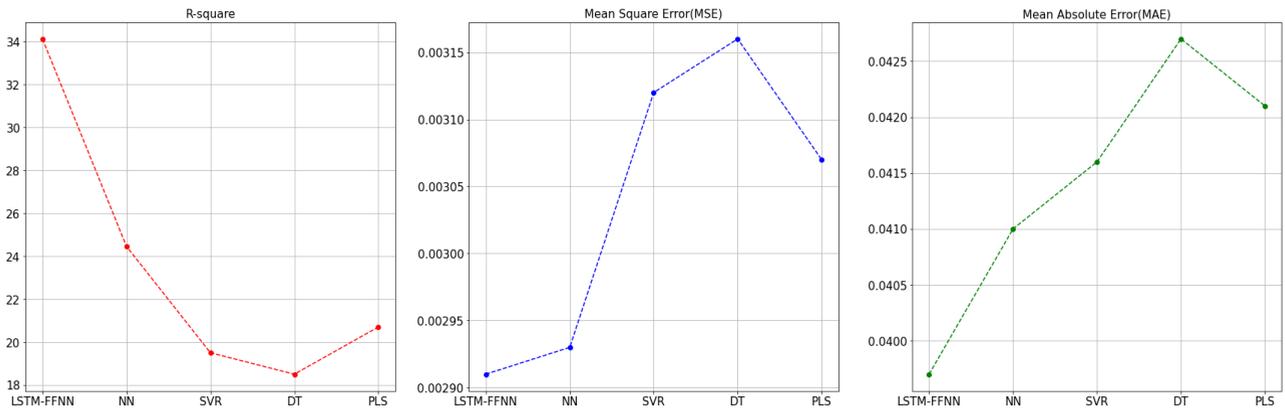


FIGURE 4. Results of the experiment.

sponds to a linear method in a high-dimensional feature space that is nonlinearly related to the input space [29].

- Decision trees [30] are one of the most widely used practical methods in statistics and machine learning in terms of both classification and regression. The target function of a decision tree has discrete output values, assigns each example to a class, and efficiently classifies new data. When the target variable takes continuous values, it is known as decision tree regression. Decision tree regression is a tree-based structure used to predict the numeric outcomes of the dependent variable, and these trees are constructed beginning with the root of the tree and proceeding down to its leaves by minimizing the predefined fitness function. The process continues until the termination criterion is satisfied.
- PLS regression [31] is a statistical regression that combines features from and generalizes principal component analysis and multiple linear regression, respectively.

The goal is to predict a set of dependent variables from a set of independent variables or predictors. This prediction is achieved by extracting from the predictors a set of orthogonal factors known as latent variables, which have the best predictive power [32].

D. EXPERIMENTAL RESULTS

The values of the three evaluation metrics of the five regression models are listed in Table 3 and illustrated in Fig. 4. As shown in the table, R^2 of the proposed combined LSTM-FFNN model is 34.14%, which is significantly higher than the R^2 values of other models. The value of R^2 for the neural network is 24.46%, indicating that the proposed method outperformed other methods. This demonstrates that the use of a sequential dataset in semiconductor manufacturing is meaningful. With respect to the MSEs and MAEs, the MSE and MAE of the combined LSTM-FFNN model are

TABLE 3. Results of the experiment.

Model	R^2	MSE	MAE
Combined LSTM–FFNN	34.13%	0.00291	0.0397
Neural Network	24.46%	0.00293	0.0410
Support Vector Regression	19.50%	0.00312	0.0416
Decision Tree	18.50%	0.00316	0.0427
PLS Regression	20.70%	0.00307	0.0421

0.00291 and 0.0397, respectively. Thus, the MSE and MAE results are the most promising.

V. CONCLUSION

In this paper, we proposed a method for edge yield prediction using a combined LSTM–FFNN model. Unlike previous research, we focused on the edge yield owing to the higher yield loss at wafer edges. Six critical process steps with four types of fabrication parameters (metrology, virtual metrology, equipment output data, and equipment name information) were used as features. The metrology, virtual metrology, and equipment output data of step A were connected via time-series to the LSTM model, and the other equipment output data and one-hot encoded equipment name information were used as inputs to the FFNN. Four regression algorithms—a neural network, support vector regression, decision tree, and PLS regression—were compared with the proposed model. The experimental results showed that the neural network outperformed the other regression models in terms of all three evaluation metrics (MSE, MAE, and coefficient of determination). The combined LSTM–FFNN model outperforms the other models with regard to all evaluation metrics. Moreover, the sequential nature of the fabrication parameters proved to be important. The following problems remain to be addressed in future research. First, additional process steps and parameters should be considered for a high-quality prediction model; this is vital because semiconductor manufacturing involves hundreds of process steps. However, both missing productivity values and time limitations present problems. Furthermore, the number of missing values may increase with the number of additional process steps or parameters. A more accurate prediction model is needed to consider the application of series metrology, a system that consistently measures the same wafers across several metrology steps. Second, in this paper, we only proposed a model for edge yield prediction. Thus, a methodological extension of total yield prediction should be researched.

REFERENCES

- [1] C. K. Shin and S. C. Park, "A machine learning approach to yield management in semiconductor manufacturing," *Int. J. Prod. Res.*, vol. 38, no. 17, pp. 4261–4271, 2000.
- [2] A. K. W. Chee, "Principles of high-resolution dopant profiling in the scanning helium ion microscope, image widths, and surface band bending," *IEEE Trans. Electron Devices*, vol. 66, no. 11, pp. 4883–4887, Nov. 2019.
- [3] A. K. W. Chee, "Quantitative dopant profiling by energy filtering in the scanning electron microscope," *IEEE Trans. Device Mater. Rel.*, vol. 16, no. 2, pp. 138–148, Jun. 2016.
- [4] P. Kang, H.-J. Lee, S. Cho, D. Kim, J. Park, C.-K. Park, and S. Doh, "A virtual metrology system for semiconductor manufacturing," *Expert Syst. Appl.*, vol. 36, no. 10, pp. 12554–12561, 2009.
- [5] A. K. W. Chee, "Unravelling new principles of site-selective doping contrast in the dual-beam focused ion beam/scanning electron microscope," *Ultramicroscopy*, vol. 213, Jun. 2020, Art. no. 112947.
- [6] J. C. Yung-Cheng and F.-T. Cheng, "Application development of virtual metrology in semiconductor industry," in *Proc. 31st Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Nov. 2005, p. 6.
- [7] P. Chen, S. Wu, J. Lin, F. Ko, H. Lo, J. Wang, C. Yu, and M. Liang, "Virtual metrology: A solution for wafer to wafer advanced process control," in *Proc. IEEE Int. Symp. Semiconductor Manuf.*, Sep. 2005, pp. 155–157.
- [8] S. J. Qin, G. Cherry, R. Good, J. Wang, and C. A. Harrison, "Semiconductor manufacturing process control and monitoring: A fab-wide framework," *J. Process Control*, vol. 16, no. 3, pp. 179–191, 2006.
- [9] N. Kumar, K. Kennedy, K. Gildersleeve, R. Abelson, C. M. Mastrangelo, and D. C. Montgomery, "A review of yield modelling techniques for semiconductor manufacturing," *Int. J. Prod. Res.*, vol. 44, no. 23, pp. 5019–5036, Dec. 2006.
- [10] International Roadmap for Devices and Systems. (2020). *Factory Integration*. [Online]. Available: <https://irids.ieee.org>
- [11] I. A. N. Goh, H. S. Chua, T. L. Neo, Y. Y. Soh, I. C. Chiang, E. W. Tan, G. Y. Tey, K. J. How, K. F. Wong, and W. Yeoh, "An integrated engineering approach to improve wafer edge yield," in *Proc. IEEE Int. Symp. Semiconductor Manuf.*, Oct. 2001, pp. 351–354.
- [12] M. Liebens, A. Jourdain, J. De Vos, T. Vandeweyer, A. Miller, E. Beyne, S. Li, G. Bast, M. Stoerring, S. Hiebert, and A. Cross, "In-line metrology for characterization and control of extreme wafer thinning of bonded wafers," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 1, pp. 54–61, Feb. 2019.
- [13] O. Yavas, E. Richter, C. Kluthe, M. Sickmoeller, and Q. AG, "Wafer-edge yield engineering in leading-edge DRAM manufacturing," *Semicond. Fabtech*, vol. 39, pp. 1–5, 2009.
- [14] T.-S. Li, C.-L. Huang, and Z.-Y. Wu, "Data mining using genetic programming for construction of a semiconductor manufacturing yield rate prediction system," *J. Intell. Manuf.*, vol. 17, no. 3, pp. 355–361, Jun. 2006.
- [15] C.-F. Chien, W.-C. Wang, and J.-C. Cheng, "Data mining for yield enhancement in semiconductor manufacturing and an empirical study," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 192–198, Jul. 2007.
- [16] W. S. Nam and S. B. Kim, "A prediction of wafer yield using product fabrication virtual metrology process parameters in semiconductor manufacturing," *J. Korean Inst. Ind. Eng.*, vol. 41, no. 6, pp. 572–578, Dec. 2015.
- [17] C.-F. Chien, C.-W. Liu, and S.-C. Chuang, "Analysing semiconductor manufacturing big data for root cause detection of excursion for yield enhancement," *Int. J. Prod. Res.*, vol. 55, no. 17, pp. 5095–5107, Sep. 2017.
- [18] S.-J. Jang, J.-S. Kim, T.-W. Kim, H.-J. Lee, and S. Ko, "A wafer map yield prediction based on machine learning for productivity enhancement," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 400–407, Nov. 2019.
- [19] D. An, H.-H. Ko, T. Gulambar, J. Kim, J.-G. Baek, and S.-S. Kim, "A semiconductor yields prediction using stepwise support vector machine," in *Proc. IEEE Int. Symp. Assem. Manuf.*, Nov. 2009, pp. 130–136.
- [20] J. Yang, S. Lee, S. Kang, S. Cho, Y. Lee, and H. Park, "Ranking process parameter association with low yield wafers using spec-out event network analysis," *Comput. Ind. Eng.*, vol. 113, pp. 419–424, Nov. 2017.
- [21] K. B. Lee, S. Cheon, and C. O. Kim, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 2, pp. 135–142, May 2017.
- [22] T. Chen, "Anomaly detection in semiconductor manufacturing through time series forecasting using neural networks," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2018.
- [23] E. Kim, S. Cho, B. Lee, and M. Cho, "Fault detection and diagnosis using self-attentive convolutional neural networks for variable-length sensor data in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 3, pp. 302–309, Aug. 2019.
- [24] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *J. Amer. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, Dec. 1952.
- [25] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [26] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.

- [27] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
- [28] C. Cortes and V. Vapnik, "Support vector machines," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [29] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 2008.
- [30] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
- [31] P. Geladi and B. R. Kowalski, "Partial least-squares regression: A tutorial," *Analytica Chim. Acta*, vol. 185, pp. 1–17, 1986.
- [32] H. Abdi, "Partial least squares regression and projection on latent structure regression (PLS regression)," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 1, pp. 97–106, 2010.



DASOL KIM is currently pursuing the M.S. degree in industrial engineering with Yonsei University. Since 2013, he has been an Engineer with the Memory Yield Enhancement Team, Samsung Electronics. His main research interests include natural language processing, machine learning, and artificial intelligence.



MINTAE KIM received the M.S. degree in industrial engineering from Yonsei University, in 2016, where he is currently pursuing the Ph.D. degree. His main research interests include natural language processing, machine learning, and artificial intelligence.



WOOJU KIM received the Ph.D. degree in operations research from KAIST, South Korea, in 1994. He is currently a Professor with the School of Industrial Engineering, Yonsei University. His main research interests include natural language processing, reliable knowledge discovery, big data intelligence, machine learning, and artificial intelligence.

...